

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-104792

(43)Date of publication of application : 21.04.1995

(51)Int.Cl. G10L 7/02

G10L 9/18

(21)Application number : 05-247184

(71)Applicant : NIPPON TELEGR & TELEPH
CORP <NTT>

(22)Date of filing : 01.10.1993

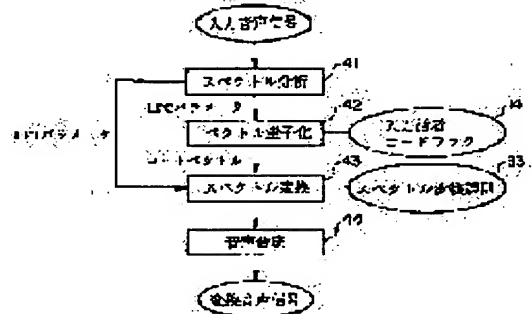
(72)Inventor : MIZUNO HIDEYUKI
ABE MASANOBU

(54) VOICE QUALITY CONVERTING METHOD

(57)Abstract:

PURPOSE: To provide a voice quality converting method which controls voice quality while maintaining the quality of a voice.

CONSTITUTION: The method consists of a step 41 wherein the spectrum of an input voice signal is analyzed, a step 42 for performing the vector quantization of LPC parameters, obtained in the step 41, on the basis of a previously generated input speaker code book, and a step 43 wherein a conversion rule corresponding to the code vector obtained in the step 42 is selected among spectrum conversion rules 33 that make 1st-4th formants F1-F4 indicating the features of speech data 13 for input speaker learning and 1st-4th formants F'1-F'4 showing the features of voice data on a speaker to be converted correspond to each other and the FFT parameter (spectrum) of the input voice signal obtained in the step 41 is converted by using the conversion rule.



LEGAL STATUS

[Date of request for examination] 15.10.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3282693

BEST AVAILABLE COPY

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-104792

(43) 公開日 平成7年(1995)4月21日

(51) Int.Cl.⁶

G 1 0 L 7/02
9/18

識別記号

D
E

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数 2 O L (全 6 頁)

(21) 出願番号 特願平5-247184

(22) 出願日 平成5年(1993)10月1日

(71) 出願人 000004226

日本電信電話株式会社
東京都千代田区内幸町一丁目1番6号

(72) 発明者 水野 秀之

東京都千代田区内幸町一丁目1番6号 日
本電信電話株式会社内

(72) 発明者 阿部 匡伸

東京都千代田区内幸町一丁目1番6号 日
本電信電話株式会社内

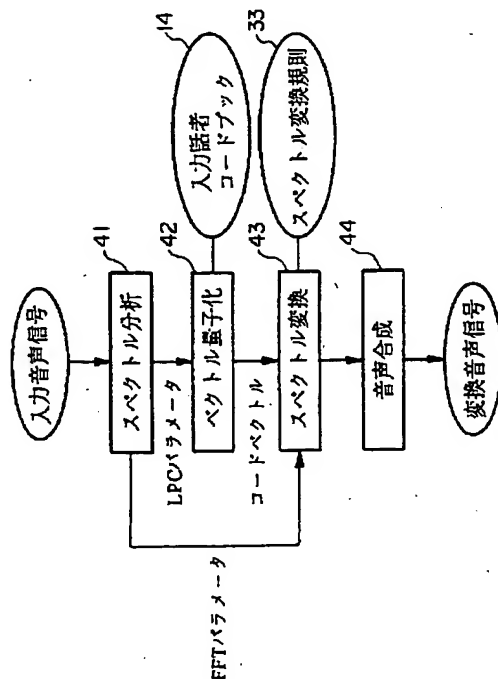
(74) 代理人 弁理士 志賀 正武

(54) 【発明の名称】 声質変換方法

(57) 【要約】

【目的】 音声の品質を保ちつつ、声質を制御する声質変換方法を提供する。

【構成】 入力音声信号をスペクトル分析するステップ41と、ステップ41で得られたLPCパラメータを、予め作成しておいた入力話者コードブック14に基づいてベクトル量子化するステップ42と、ステップ42で得られたコードベクトルに対応する変換規則を、入力話者学習用音声データ13の特徴を示す第1～第4のフォルマントF1～F4と変換対象話者音声データ23の特徴を示す第1～第4のフォルマントF'1～F'4とを対応付けたスペクトル変換規則33から選択し、この変換規則を用いて、ステップ41で得られた入力音声信号のFFTパラメータ(スペクトル)を変換するステップ43とからなる。



【特許請求の範囲】

【請求項1】 入力話者による入力音声を、前記入力話者と異なる変換対象話者の声質を有する音声に変換する声質変換方法において、

前記入力音声の波形をスペクトル分析するスペクトル分析過程と、

前記スペクトル分析過程で得られた分析結果を、予め作成しておいた入力話者のコードブックに基づいてベクトル量子化するベクトル量子化過程と、

前記ベクトル量子化過程で得られたコードベクトルに対応する変換規則を、前記入力音声の特徴と前記変換対象話者の音声の特徴とを統計的な手法を用いて対応付けたスペクトル変換規則から選択し、この変換規則を用いて、前記スペクトル分析過程で得られた前記入力音声の波形のスペクトルを変換するスペクトル変換過程とからなり、

前記スペクトル変換過程で変換されたスペクトルに応じた音声出力されることを特徴とする声質変換方法。

【請求項2】 前記入力話者のコードブックは、前記入力話者による学習用音声の波形をスペクトル分析し、その分析結果であるコードベクトルを統計的に分類してなり、

前記スペクトル変換規則は、

前記入力話者コードブックを用いて表現されるスペクトル特徴量と、前記学習用音声と発声内容同一の学習用音声を前記変換対象話者に発声させ、この学習用音声の波形をスペクトル分析し、その分析結果をベクトル量子化することにより得られた変換対象話者のコードブックとを順次対応付け、

前記スペクトル特徴量と、

前記対応付けの結果を用いて、前記入力話者のコードブックのコードベクトル毎に前記変換対象話者コードブックを平均化して得られたマッピングコードブックで表現されるスペクトル特徴量とを対応付けてなる変換規則から構成されることを特徴とする請求項1に記載の声質変換方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、入力話者の音声を、所望する話者の声質を持つ音声に変換する声質変換方法に関する。

【0002】

【従来の技術】 従来より、音声の声質変換方法として、線形予測分析・合成方式（以後、LPC（Linear Predictive Coding）分析・合成方式と称す）に基づいて、音声スペクトル包絡特性を表す各種パラメータを算出し、これらのパラメータを変更することにより音声の声質を変更する方法や、変換元の話者（以後、入力話者と称す）と変換先の話者（以後、変換対象話者と称す）との音声波形またはスペクトルの対応関係を予め求めてお

き、その対応関係に従って、入力話者によって発生された音声を、変換対象話者の音声へ変換する方法等が知られている。

【0003】 ここでは、LPC分析・合成方式に基づく声質変換方法の概略を説明する。従来のLPC分析・合成方式に基づく方法では、声帯から唇までの声道の特徴を表す線形予測係数（以後、LPCパラメータと称す）、音源（声帯の振動）を表すパルスやローセンバーク波等のパラメータを、入力話者および変換対象話者について採取し、両者間の各パラメータの対応関係を、適当なサンプルデータから実験的または経験的に把握して声質の変換規則を決定する。

【0004】 そして、入力話者の入力音声を変換する際には、入力音声信号から上記各パラメータを算出し、予め決定された上記変換規則に従って各パラメータを変換し、再合成することにより、出力される音声の声質を、変換対象話者のものに変換する。上述したLPC分析・合成方式に基づく声質変換方法の詳細は、例えば、D. G. CHILDERS and Ke WU, "VOICE CONVERSION" (Speech Communication 8 (1989) pp.147-158) に記載されている。

【0005】

【発明が解決しようとする課題】 ところで、上述した従来の声質変換方法に用いられる変換規則は、適当なサンプルデータから実験的または経験的に決定されたものであるため、入力話者が発する任意の入力音声を適切に変換できるという保証は無い。

【0006】 また、入力話者が実際に発する音声において、LPCパラメータと音源パルスを表すパラメータ間には複雑な相関関係があり、それらを全て考慮した変換規則を決定することは極めて困難である。このため、従来の声質変換方法を用いて声質変換を行った場合、変換された音声において、音韻が変化してしまう等の品質劣化が発生することがあるという問題があった。本発明は、上述した事情に鑑みて為されたものであり、音声の品質を保ちつつ、声質を制御する声質変換方法を提供することを目的とする。

【0007】

【課題を解決するための手段】 本発明による声質変換方法は、入力話者による入力音声を、前記入力話者と異なる変換対象話者の声質を有する音声に変換する声質変換方法において、前記入力音声の波形をスペクトル分析するスペクトル分析過程と、前記スペクトル分析過程で得られた分析結果を、予め作成しておいた入力話者のコードブックに基づいてベクトル量子化するベクトル量子化過程と、前記ベクトル量子化過程で得られたコードベクトルに対応する変換規則を、前記入力音声の特徴と前記変換対象話者の音声の特徴とを統計的な手法を用いて対応付けたスペクトル変換規則から選択し、この変換規則を用いて、前記スペクトル分析過程で得られた前記入力

音声の波形のスペクトルを変換するスペクトル変換過程とからなり、前記スペクトル変換過程で変換されたスペクトルに応じた音声出力されることを特徴としている。

【0008】

【作用】上記方法によれば、スペクトル分析の結果は、入力話者のコードブックに基づいてベクトル量子化され、このベクトル量子化で得られたコードベクトルに対応する変換規則が、スペクトル変換規則から選択されて入力音声の波形に適用される。前記変換規則は、入力音声の特徴と変換対象話者の音声の特徴とを統計的な手法を用いて対応付けたものであり、入力音声に対して適応的に選択される。このため、音声の品質を保ちつつ、声質を制御することが可能となる。

【0009】

【実施例】以下、図面を参照して、本発明の一実施例について説明する。図1(a)は、本発明の一実施例による声質変換方法の一部手順を示すフローチャートである。この図に示す手順は、音声信号を効率良く表現するために、音声信号の特徴を示すパラメータ（以後、音声特徴量と称す）を算出し、算出された音声特徴量を統計的に分類し、コードブックと呼ばれる分類表を作成するというものである。なお、音声特徴量としては、LPC分析によるLPCパラメータやFFT（fast Fourier transform）分析によるスペクトル密度等があるが、ここではLPCパラメータを用いた例について説明する。

【0010】図1(a)において、まず、ステップ11では、入力話者により発生された入力音声に対応する入力話者学習用音声データ13に対して、前述したLPC分析処理が施され、LPCパラメータが算出される。LPC分析は、統計的に正確さを期すため、充分に多くの入力話者学習用音声データ13に対して施される。次に、ステップ12では、収集されたLPCパラメータに対して、クラスタリング（分類）が行われる。クラスタリングの手法としては、代表的な方法であるLBG（Linde-Buzo-Gray）アルゴリズム等がある。LBGアルゴリズムの詳細は、例えば、Lindeら、「An algorithm for Vector Quantization Design」（IEEECOM-28(1980-01)）に記載されている。

【0011】上述した手順を経て、入力話者コードブック14が作成される。図1(b)は入力話者コードブック14の構成を示す概念図であり、この図に示すように、入力話者コードブック14は、通常256～512程度のコードベクトル15から構成される。各コードベクトル15において、16はコードベクトル番号であり、例えば、1～256の自然数が順に割り当てられる。17は入力話者学習用音声データ13に対応するスペクトル特徴量であり、ここでは、数個のLPCパラメータで構成されている。

【0012】次に、スペクトル変換規則を決定する際に

使用されるマッピングコードブック28を作成する過程を、図2を参照して説明する。マッピングコードブック28は、入力話者の音声信号と変換対象話者の音声信号とを統計的に対応付けるものである。まず、ステップ21において、変換対象話者学習用音声データ23から変換対象話者コードブック22が作成される。この作成手順は、図1(a)に示す手順と同一であるので、その説明を省略する。

【0013】次に、ステップ24、24では、入力話者および変換対象話者コードブック14、22に基づいて、入力話者学習用音声データ13および変換対象話者学習用音声データ23それぞれに、LPC分析およびベクトル量子化処理が施される。ここで、ベクトル量子化処理とは、各音声データ13、23をLPC分析して得られたLPCパラメータに、最も似かよっているスペクトル特徴量17を有するコードベクトル15を各コードブック14、22から抽出し、抽出されたコードベクトル15中のスペクトル特徴量17を出力する処理である。ベクトル量子化の詳細については、例えば、古井貞照著、「デジタル音声処理」に記載されている。

【0014】上述したベクトル量子化処理により、変換対象話者コードベクトル系列25および入力話者コードベクトル系列26が得られる。次に、ステップ27では、入力話者コードベクトル系列26および変換対象話者コードベクトル系列25から、両者を対応付けるマッピングコードベクトルが生成される。マッピングコードベクトルからマッピングコードブック28が作成される。

【0015】マッピングコードベクトルの生成方法としては、各入力話者コードベクトル系列26毎に、対応する複数の変換対象話者コードベクトル系列25を集計し、重み付け平均化により生成する公知の方法を用いる。この方法の詳細は、例えば、阿部ら、「Voice Conversion through vector quantization」（JASJ(E) 11, 2 (1990) pp. 71-76）に記載されている。

【0016】こうして作成されたマッピングコードブック28を用いて、スペクトル変換規則33を作成する過程を、図3を参照して説明する。スペクトル変換規則33は、音声の個人性に関する特徴量の一つであるフォルマント周波数を変換する規則である。図3において、まず、ステップ31、31では、入力話者コードブック14中の各コードベクトル15と、マッピングコードブック28中の各マッピングコードベクトルとに、それぞれフォルマント分析を施す。これにより、各ベクトルに対するフォルマント周波数が求められる。

【0017】フォルマント周波数の分析手法は多々あり、例えば、LPC極抽出に基づく方法を簡便に用いることができる。フォルマント周波数の分析手法の詳細は、例えば、板倉ら、「統計的手法による音声スペクトル密度とホルマント周波数の推定」（信学論、(1970). 5

5

3-A, 1, pp. 35-42) に記載されている。

【0018】次に、ステップ32では、スペクトル変換規則33が求められる。具体的には、まず、図4に示すように、入力話者コードブック14中のコードベクトル15における第1～第4フォルマントF1～F4を求める。次に、このコードベクトル15に対応するマッピングコードベクトルを、マッピングコードブック28から検索し、当該マッピングコードベクトルから変換対象話者に対応するコードベクトルを抽出する。そして、抽出されたコードベクトルにおける第1～第4フォルマントF'1～F'4を求め、それぞれ、上記第1～第4のフォルマントF1～F4と対応させる。両者の対応付けは、自動的に、あるいは手作業で行われる。

【0019】次に、第1～第4のフォルマントF1～F4に対応する周波数 ω_1 , ω_2 , ω_3 , ω_4 と、第1～第4のフォルマントF'1～F'4に対応する周波数 ω'_1 , ω'_2 , ω'_3 , ω'_4 をスペクトル変換規則33に記録する。ここで、音韻種別によっては第4フォルマントが存在しない場合があり、その際は、第4フォルマントについては記録しない。

【0020】こうして、スペクトル変換規則33が作成される。スペクトル変換規則33の一例を図5に示す。この図に示すように、スペクトル変換規則33は、複数のレコードから構成され、各レコードには、1～256の自然数であるスペクトル変換規則番号34が割り当てられている。このスペクトル変換規則番号34は、入力話者コードブック14中のコードベクトル番号16と1対1で対応するように割り当てられている。

【0021】また、各レコードには、第1～第4のフォルマント別に、対応する周波数が記録されている。例えば、スペクトル変換規則番号が「1」であるレコードでは、第1のフォルマントについて、周波数 ω_1 (710)と周波数 ω'_1 (815)とが対応付けられて記録されている。

【0022】上述した過程を経て作成された、スペクトル変換規則33を用いて、入力音声信号を声質の異なる変換音声信号に変換する過程を、図6を参照して説明する。図6において、まず、ステップ41では、入力音声信号に対して、スペクトル分析処理が行われる。スペクトル分析処理は、LPC分析処理およびFFT分析処理からなり、入力音声信号に応じたLPCパラメータおよびFFTパラメータ(スペクトル)が得られる。

【0023】次に、ステップ42では、ステップ41で得られたLPCパラメータを、予め作成しておいた入力話者コードブック14に基づいてベクトル量子化する。これにより、入力音声信号に対応するコードベクトルが得られる。次に、ステップ43では、ステップ41で得られたFFTパラメータを変換する。この変換過程を以下に説明する。

【0024】具体的には、まず、ステップ42で得られ

6

たコードベクトルに対応するレコードを、予め作成しておいたスペクトル変換規則33から抽出する。そして、抽出されたレコードに表される変換規則に従って、ステップ41で得られたFFTパラメータ(スペクトル)のフォルマント周波数を変換する。フォルマント周波数の変換方法の詳細は、水野ら、「制御自由度の高いフォルマント周波数変換法」(音講論集, pp. 319-340)に記載されているため、ここでは、その概略を説明するにとどめる。

10 【0025】本実施例の変換方法では、入力音声信号を1ピッチ単位で切り出し、LPC極分析によって入力音声のフォルマントを抽出する。そして、あるフォルマントの周波数を変換する際には、当該フォルマントのスペクトル密度と、当該フォルマントにおいて所望するスペクトル密度との差を、繰り返し処理によって一定値以下に抑えつつ、所望するフォルマント周波数が変換された全極型スペクトル特性を決定する。次に、こうして得られた全極型スペクトル特性を有する全極型フィルタを構成し、所望するフォルマント周波数特性が得られるまで

20 繰り返し原音声に対して作用させて、所望するフォルマント周波数に音声を変換する。

【0026】次に、ステップ44では、ステップ43でスペクトル変換されて得られるFFTパラメータ(スペクトル)から、音声信号をIFFTにより合成し、変換音声信号を出力する。この変換音声信号は、変換対象話者の声質を有したものとなる。

【0027】以上説明したように、入力話者コードブック14中のコードベクトル15における第1～第4フォルマントF1～F4と、このコードベクトル15に対応するマッピングコードベクトルにおける第1～第4フォルマントF'1～F'4とが対応付けられている。また、上記マッピングコードベクトルは、入力話者コードブック14中の各コードベクトル15に対応して重み付け平均化された変換対象話者コードブック22から生成されている。したがって、上記スペクトル変換規則33を用いることにより、入力音声に対して適応的な変換を行うことができる。これにより、変換音声信号は高品質なものとなることが保証される。

【0028】

40 【発明の効果】以上説明したように、本発明によれば、スペクトル分析の結果は、入力話者のコードブックに基づいてベクトル量子化され、このベクトル量子化で得られたコードベクトルに対応する変換規則が、スペクトル変換規則から選択されて入力音声の波形に適用される。前記変換規則は、入力音声の特徴と変換対象話者の音声の特徴とを統計的な手法を用いて対応付けたものであり、入力音声に対して適応的に選択される。したがって、音声の品質を保ちつつ、声質を制御することができるという効果がある。

【図面の簡単な説明】

【図1】本発明の一実施例による声質変換方法を説明するための図である。

【図2】マッピングコードブック28の作成過程を示す図である。

【図3】スペクトル変換規則33の作成過程を示す図である。

【図4】スペクトル変換規則33を説明するための図である。

【図5】スペクトル変換規則33の構成を示す概念図で

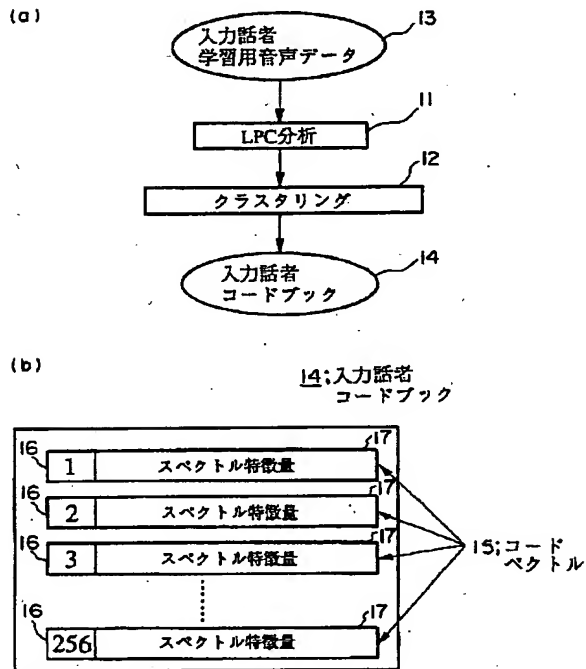
ある。

【図6】スペクトル変換規則33を用いた声質変換過程を示す図である。

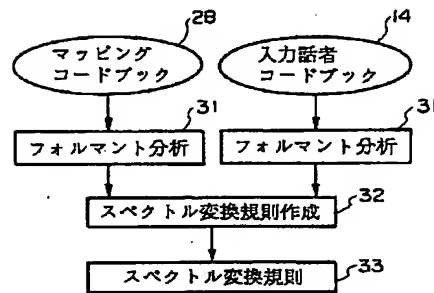
【符号の説明】

- 14 入力話者コードブック
- 22 変換対象話者コードブック
- 28 マッピングコードブック
- 33 スペクトル変換規則

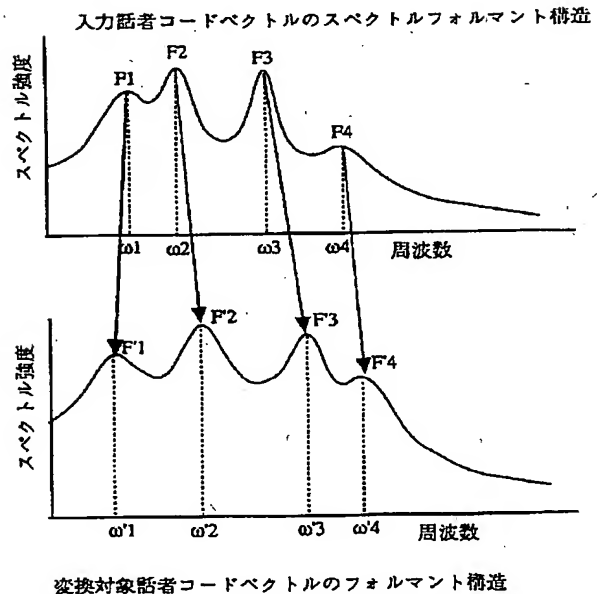
【図1】



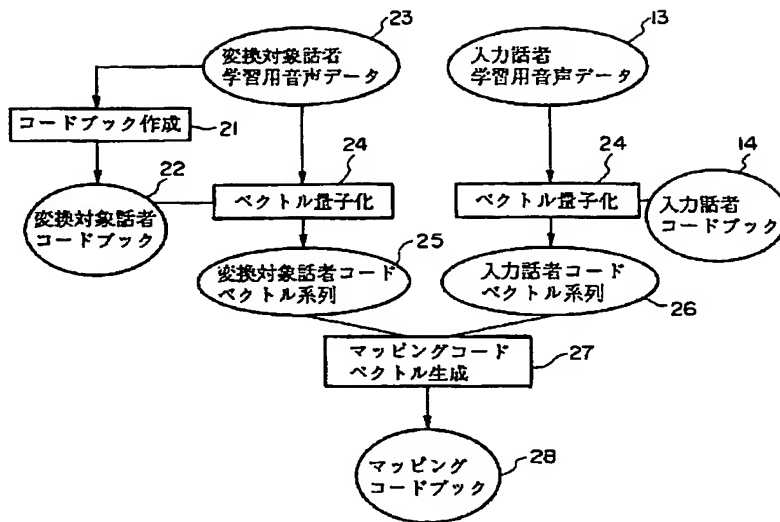
【図3】



【図4】



【図 2】

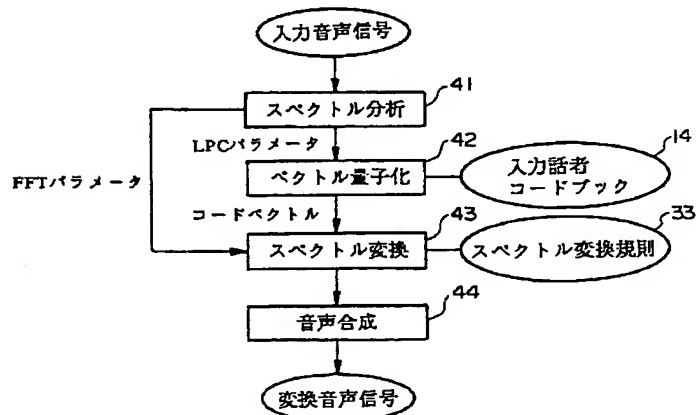


【図 5】

34; スペクトル変換規則番号

	F1	ω_1	ω'_1	F2	ω_2	ω'_2	F3	ω_3	ω'_3	F4	ω_4	ω'_4
1	F1	710	815	F2	1480	1510	F3	2120	2030	F4	2940	3100
2	F1	670	805	F2	1495	1500	F3	2100	2230	F4	3040	3120
3	F1	680	815	F2	1475	1590	F3	2010	1990	F4	2980	3210
256	F1	420	605	F2	980	1021	F3	1820	1730	F4	2540	2190

【図 6】



BEST AVAILABLE COPY